ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

# In Vitro Reconstitution of Mevalonate Pathway and Targeted Engineering of Farnesene Overproduction in *Escherichia coli*

**Fayin Zhu,[1] Xiaofang Zhong,[1] Mengzhu Hu,[1] Lei Lu,[1] Zixin Deng,[1,2] Tiangang Liu[1,2]**

[1]Key Laboratory of Combinatorial Biosynthesis and Drug Discovery (Wuhan University), Ministry of Education, and Wuhan University School of Pharmaceutical Sciences, Wuhan 430071, PR China; telephone: +86-27-68755086; fax: +86-27-68755086; e-mail: liutg@whu.edu.cn

[2]Hubei Engineering Laboratory for Synthetic Microbiology, Wuhan Institute of Biotechnology, Wuhan 430075, PR China

**ABSTRACT:** Approaches using metabolic engineering and synthetic biology to overproduce terpenoids, such as the precursors of taxol and artemisinin, in microbial systems have achieved initial success. However, due to the lack of steady-state kinetic information and incomplete understanding of the terpenoid biosynthetic pathway, it has been difficult to build a highly efficient, universal system. Here, we reconstituted the mevalonate pathway to produce farnesene (a precursor of new jet fuel) in vitro using purified protein components. The information from this in vitro reconstituted system guided us to rationally optimize farnesene production in *E. coli* by quantitatively overexpressing each component. Targeted proteomic assays and intermediate assays were used to determine the metabolic status of each mutant. Through targeted engineering, farnesene production could be increased predictably step by step, up to 1.1 g/L (~2,000 fold) 96 h after induction at the shake-flask scale. The strategy developed to release the potential of the mevalonate pathway for terpenoid overproduction should also work in other multistep synthetic pathways.

Biotechnol. Bioeng. 2014;9999: 1–10.



**KEYWORDS:** mevalonate pathway; in vitro reconstitution; terpenoid; farnesene; targeted engineering; biofuel

Correspondence to: T. Liu

Contract grant sponsor: Ministry of Science and Technology of China 973
Grant numbers: 2011CBA00800; 2012CB721000
Contract grant sponsor: Ministry of Science and Technology of China 863
Contract grant number: 2012AA02A701
Contract grant sponsor: National Natural Science Foundation of China
Contract grant number: 31222002
Contract grant sponsor: New Century Excellent Talents in University
Contract grant sponsor: Ministry of Education of China
Contract grant number: 20110141120028
Received 10 December 2013; Revision received 17 January 2014; Accepted 21 January 2014
Accepted manuscript online xx Month 2014;
Article first published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/bit.25198

## Introduction

Terpenoids belong to the most diverse family of natural products, containing more than 25,000 structures elucidated from plants, microorganisms, insects, and various marine organisms. Two distinct pathways, the mevalonate (MVA) pathway and the methylerthritol-phosphate (MEP) pathway (Phillips et al., 2008; Rohmer et al., 1993), are responsible for the synthesis of the two universal 5-carbon isoprenoid building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Downstream of both pathways contain kinds of enzymes which catalyze the linear condensation and eventual cyclization of the individual building blocks, as well as tailoring enzymes, including various oxygenases, for generating the final terpenoid products (Kirby and Keasling, 2009). Terpenoid metabolites have many applications, including medicinal (taxol and artemisinin), nutraceutical (lycopene and carotenoids), industrial (isoprene), and even as precursors of next-generation jet fuel (farnesene) (Fig. 1) (Ajikumar et al., 2010; Farmer and Liao, 2000; Martin et al., 2003). Whether for drugs or biofuels, the low levels of productivity and high cost of terpenoid products have significantly hindered their application. Recent developments in metabolic engineering and synthetic biology have allowed overproduction of terpenoids based on microbial fermentation rather than plant-based production, resulting in several remarkable breakthroughs, not only in the production of complex natural products, such as precursors of taxol and artemisinin (Ajikumar et al., 2010; Paddon et al., 2013; Westfall et al., 2012), but also in the production of bulk chemicals and biofuels (Ajikumar et al., 2008; Alper et al., 2005; Farmer and Liao, 2000; Zhang et al., 2011). The successful overproduction of these compounds, especially the amorphadiene and taxadiene, has encouraged the hope that any terpenoid

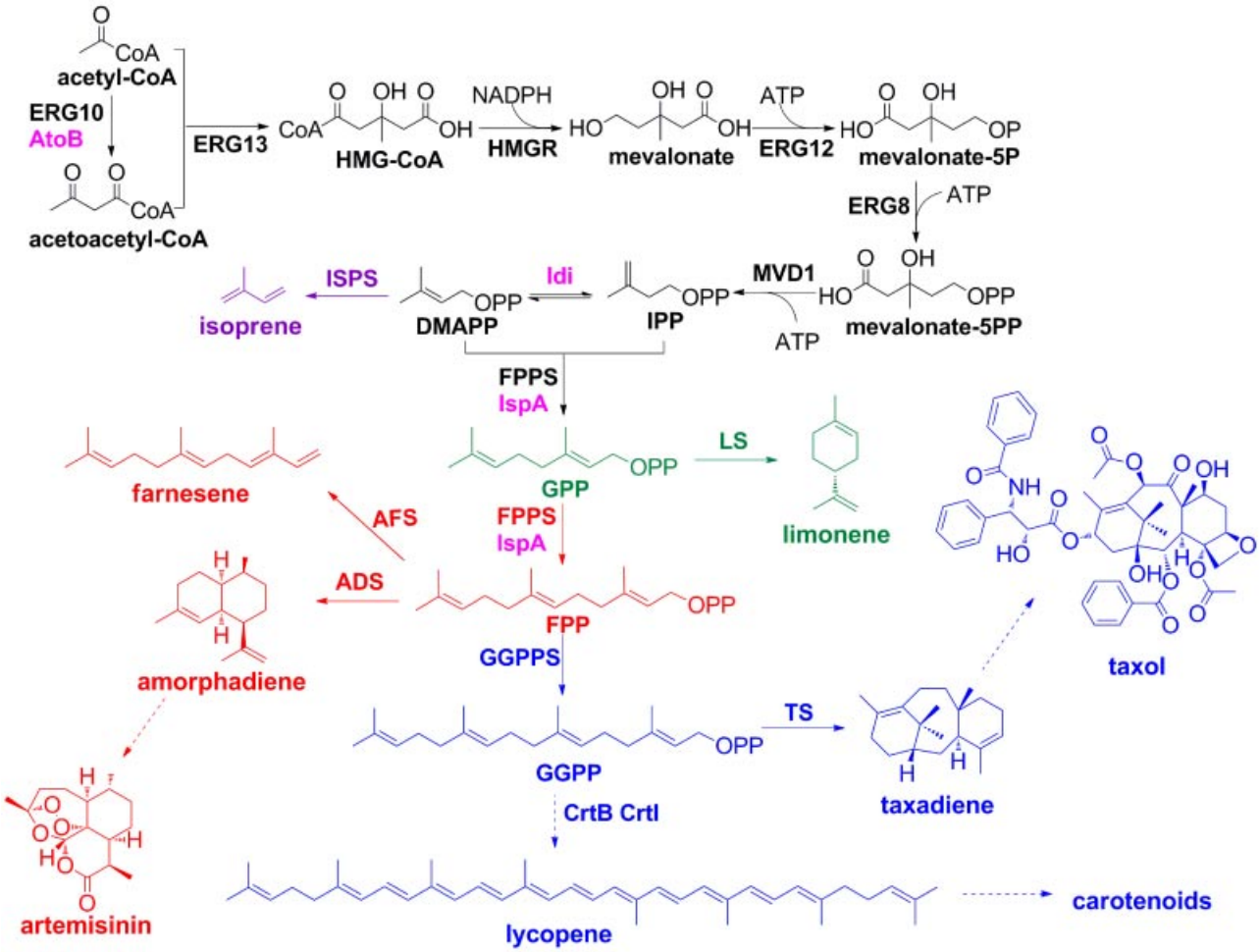


**Figure 1.** Mevalonate pathway for terpenoid biosynthesis: ERG10, acetyl-CoA acetyltransferase from *S. cerevisiae*; AtoB, acetyl-CoA acetyltransferase from *E. coli*; ERG13, 3-hydroxy-3-methylglutarylcoenzyme A (HMG-CoA) synthase; HMGR, HMG-CoA reductase; ERG12, mevalonate kinase; ERG8, phosphomevalonate kinase; MVD1, mevalonate pyrophosphate decarboxylase; Idi, isopentenyldiphosphate (IPP) isomerase; ISPS, isoprene synthase; FPPS, farnesyl pyrophosphate (FPP) synthase; IspA, FPPS from *E. coli*; LS, limonene synthase; AFS, α-farnesene synthase; ADS, amorphadiene synthase; GGPPS, geranylgeranyl pyrophosphate (GGPP) synthase; TS, taxadiene synthase; CrtB, phytoene synthase; CrtI, phytoene desaturase; DMAPP, dimethylallyl pyrophosphate; GPP, geranyl pyrophosphate; CrtE, a form of GGPPS, not shown here because it can condense IPP and DMAPP directly to form GGPP.

metabolic pathway can be rationally engineered for enhanced productivity.

Despite these advances, there are still no guiding principles available for the construction of new, highly efficient terpenoid pathways in microbial hosts due to the lack of both steady-state kinetics information and detailed knowledge of the factors that influence the metabolic flux through terpenoid biosynthetic pathways. The successes to date were achieved by using combinatorial approaches involving the construction of numerous mutants. Although many high-throughput screening methods have recently been developed for rapid analysis of various combinations of each enzyme component under the control of different promoters in either *E. coli* or yeast (Gibson, 2011; Shao and Zhao, 2009), it is impossible to obtain the ideal combination purely by random selection and there is no fast and convenient analytical method to screen and evaluate the mutants in most cases. Despite the successful examples of amorphadiene and taxadiene production, how to optimization of the system remains a challenge. How much room remains for improvement or what is the optimum direction for further engineering.

As an alternative strategy, we had established an in vitro reconstituted system for *E coli* fatty acid synthase that allows monitoring of the steady-state kinetic and biochemical parameters and the accumulation of intermediates. The information thus gained could be used to guide the optimization of each biosynthetic component in *E. coli* for further improvement of the production of fatty acid-derived compounds in vivo (Yu et al., 2011). Therefore, we sought to build on this strategy by undertaking a systematic kinetic analysis of a fully reconstituted terpenoid pathway so as to unveil the mechanism of terpenoid synthesis, to assess the contribution of each enzyme and cofactors, to identify the key enzymes and rate-limiting step in an ideal biosynthetic production system.

The information gained from the in vitro system could then guide the engineering of an optimized system in vivo. Additional control was achieved by precise measurement of the relative levels of protein and intermediate in engineered mutant hosts using a modified targeted proteomics method (Redding-Johanson et al., 2011; Singh et al., 2012) and mass spectrometry (MS)-based intermediate analysis. Compared to traditional metabolic engineering methods, each stage in the evolution of the biosynthetic pathway can be rationally controlled and precisely evaluated against objectively determined rather than empirically chosen milestones (Fig. S1). Because isopentenyl diphosphate isomerase (Idi) was identified as the key enzyme of mevalonate pathway, we overexpressed an extra Idi in vivo and demonstrated a 5.5-fold increase over the control strain. This increase resulted in up to 731 mg/L farnesene produced 48 h after induction, ~2-fold higher than previously reported under the same conditions, and further increased to 1.1 g/L at 96 h after induction.

## Materials and Methods

### Plasmid Construction

Detailed information on plasmid construction is shown in the SI text. Primers are shown in Table SI and plasmids are summarized in Table SII.

### Plasmids for Protein Overexpression and Purification

The α-farnesene synthase gene from *Malus x domestica* (Pechous and Whitaker, 2004) was codon optimized and synthesized by Genescript (DNA sequence is shown in Table SV). The acetoacetyl-CoA thiolase (*atoB*), *idi*, and farnesyl pyrophosphate synthase (*ispA*) genes from *E. coli*, the 3-hydroxy-3-methylglutaryl-CoA synthase (*erg13* or *hmgs*), a truncated version of 3-hydroxy-3-methylglutaryl-CoA reductase (*thmg1*) (Polakowski et al., 1998), mevalonate kinase (*erg12*), phosphomevalonate kinase (*erg8*), and mevalonate pyrophosphate decarboxylase (*mvd1*, also known as *erg19*) genes from *Saccharomyces cerevisiae* were amplified from genomic DNA or synthesized DNA and cloned into pET28a(+) for protein overexpression and purification. An *alf* gene was also cloned into pET28a(+) for chimeric protein ALF overexpression, in which IspA and AFS were fused with a $(GGGS)_2$ linker as previously described (Wang et al., 2011).

### Plasmids for Mevalonate Pathway Gene Overexpression

Two plasmids, pMH1 and pFZ81, were constructed for overexpressing the entire set of MVA pathway genes by using simple cloning (You et al., 2012) and the Gibson method (Gibson, 2011) which can assemble multiple genes dependent on the overlap of DNA. pMH1 contains the genes encoding the first three enzymes of the mevalonate pathway, AtoB, ERG13, and tHMG1; pFZ81 contains the other four genes in the mevalonate pathway, ERG12, ERG8, MVD1, and IDI. pMH1 and pFZ81 both use the single *lac* promoter in the pBBR1MCS backbone to control expression of the indicated three or four genes.

### Plasmids for Farnesene Overproduction

Plasmid pFZ38 contains separate *ispA* and *afs* genes and was constructed for farnesene production by standard methods. The derivative plasmids from pFZ38 were similarly constructed using the isocaudarner *Spe*I and *Xba*I to splice together corresponding genes. An additional copy of *erg8*, *erg13*, or *idi* was introduced into pFZ38 by inserting the *Xba*I-*Xho*I fragment of *erg8*, *erg13*, or *idi* into the *Spe*I-*Xho*I fragment of pFZ38 to give pFZ65, pFZ69, and pFZ71, respectively. These genes were controlled by the single T7 promoter from pET21a(+).

### Protein Purification

Proteins for farnesene production via the MVA pathway were individually overexpressed and purified from recombinant strains of *E. coli* BL21(DE3) harboring the relevant plasmid. Several single colonies were incubated in 10 mL LB medium with 50 μg/mL kanamycin at 30°C until the $OD_{600}$ reached 0.6–0.8. The culture was cooled to 18°C and then 0.1 mM IPTG was added for induction. After further growth at 18°C for 12–16 h, cells were harvested by centrifugation at 6,000*g*, and resuspended in 35 mL buffer A (50 mM Tris–Cl, 300 mM NaCl, 4 mM β-mercaptoethanol, pH 7.6). The cell suspensions were lysed by sonication and centrifuged at 45,000*g* for 1 h. After filtration of the supernatant, 6% buffer B (500 mM imidazole in buffer A) was added to adjust the supernatant to 30 mM imidazole to reduce nonspecific binding. The following steps were performed on an ÄKTA Explorer 10S (GE Healthcare, Connecticut): The Histap QHP 5 mL column (GE Healthcare) was equilibrated with 30 mM imidazole (6% buffer B) for 6 CVs (Column Volumes), and then the sample was injected into the column using a sample pump. Unbound protein was washed out with 2 CVs of buffer A and then 8% buffer B (9CVs) was used to wash any nonspecific binding protein. Buffer B was linearly increased to 60% in 12 CVs, then to 100% for an additional 4 CVs, with collection of 5-mL fractions with the Frac 950 collector. The column was then re-equilibrated with 6% buffer B. The fractions containing purified target protein were concentrated using Amicon Ultra-15 centrifugal filter devices (Millipore, Darmstadt, Germany). Concentrated protein (2.5 mL) was applied to a PD-10 column (GE Healthcare) to change buffer, which was equilibrated with storage buffer (100 mM Phosphate buffer, 10% glycerol, pH 7.4). Protein concentrations were measured with a Pierce® BCA protein assay kit (Thermo Fisher Scientific, Waltham, MA) using a microplate reader. Proteins were stored at −80°C after flash freezing in liquid nitrogen.

### Reconstitution and Kinetic Analysis of Farnesene Synthetic Components

Production of farnesene in the reconstituted system was confirmed by gas chromatography–mass spectrometry (GC–MS) and radioactive TLC. In vitro assays of the reconstitution

system were performed as previously described (Yu et al., 2011). For the demonstration assay, 1 μM of each protein component (while tHMG1 at 3 μM and IspA at 2 μM), 1.2 mM NADPH, and 4 mM ATP were added to a reaction buffer containing 50 mM sodium phosphate buffer (pH 7.4), 30 mM potassium chloride, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 8 mM magnesium chloride, 0.2 mM manganese chloride, and 0.01 mM zinc sulfate. The reactions were initiated by addition of 4 mM acetyl-CoA, and then incubated at room temperature for 30 or 120 min. To confirm farnesene had accumulated in the 100 μL assay volume, the reaction mixture was extracted with 100 μL hexane. After thorough mixing, the samples were centrifuged at 12,000 g for 5 min, and then the hexane layer was transferred for GC–MS analysis. Farnesene accumulation was further confirmed by radioactive assay in which the reactions were initiated by addition of 600 μM of acetyl-CoA (10% or 5% [1-$^{14}$C] acetyl-CoA, American Radiolabeled Chemicals). After incubation at room temperature for 5, 10, 15, 30, or 45 min, farnesene in 30-μL of reaction solution was extracted with 400 μL hexane. After thorough mixing and centrifugation at 12,000 g for 5 min, 350 μL of the hexane layer was transferred to a separate tube, and dried using a Speedvac. Dried samples were resuspended in 20 μL hexane, spotted onto a silica gel TLC plate, and chromatographed using a solvent system comprised of a 70:30:2 (v/v) mixture of hexane:ether:acetic acid. Radioactivity was quantified on a Phosphorimager Storm 825 (GE Healthcare).

For steady state analysis, the protein components, substrates, cofactors, and metal ions were added to a reaction buffer containing 50 mM sodium phosphate buffer (pH 7.4) and 1 mM TCEP at the indicated concentrations. Reactions were initiated by addition of acetyl-CoA (5% [1-$^{14}$C] acetyl-CoA, American Radiolabeled Chemicals) at the indicated concentration and allowed to proceed for the designated time at room temperature. To measure the amount of farnesene that had accumulated in a 20-μL assay volume, the reaction mixture was quenched by addition to a mixture of 480 μL isopropanol:acetic acid:water with a ratio of 20:5:23 (v/v), and then extracted with 500 μL of hexane. After thorough mixing, the samples were centrifuged to expedite phase separation. Because there is only one band as farnesene on the TLC plate in the radioactive assay, the amount of farnesene produced in the reconstituted system was measured by liquid scintillation analyzer. 400 μL of the hexane layer was transferred to a separate 2 mL tube, to which was added 1 mL liquid scintillation cocktail Ultima Gold solution (PerkinElmer, Waltham, MA). After thorough mixing, the sample was analyzed with a PerkinElmer Tri-Carb liquid scintillation analyzer 2910 TR, using [1-$^{14}$C] acetyl-CoA as calibration standard. Origin v8.0 software was used for data fitting.

### Shake-Flask Fermentation and Analysis of α-Farnesene Production in Engineered Strains

To assess farnesene production in shake flasks,*E. coli* BL21 (DE3)/pFZ35 (control strain F0) and the engineered strains were cultured in 500-mL flasks containing 150 mL 2 × TY medium at 30°C with 2% glycerol, 100 μg/mL ampicillin, 50 μg/mL kanamycin, and 34 μg/mL chloramphenicol. When the $OD_{600}$ reached 0.6–0.8, IPTG was added to a final concentration of 0.1 mM. After 3 h induction, the medium was overlaid with 30 mL decane. Samples of 1 mL decane and 5 mL culture were withdrawn at designated intervals. After centrifugation, the decane phase was subjected to GC–MS analysis. The production of α-farnesene was quantified using an Agilent Technologies 7890A Gas Chromatograph equipped with a 5975MS; trans-β-farnesene in decane was used as calibration standard as described previously (Wang et al., 2011).

### Modified Targeted Proteomic Analysis in Engineered Strains

Cells were cultured as described for shake-flask fermentation and harvested 20 h after induction. A modified targeted proteomic analysis was then performed, similar to the methods described by Redding-Johanson et al. (2011). First, 100 μg of total protein was digested with trypsin and an AB SCIEX TripleTOF 5600+ was used for targeted peptide identification running with ProteinPilot 4.5. An AB SCIEX Q-Trap 4500 mass spectrometer coupled with an Eksigent NanoLC-Ultra 1Dplus system and cHiPLC-nanoflex system running with Analyst 1.6 was used for quantification. MultiQuant$^{TM}$ version 2.1 was used for data analysis. In the quantification section, Skyline 1.4 was used for the experimental designation. Multiple reaction monitoring-initiated detection and sequencing (MIDAS) (Unwin et al., 2009) was applied for the transition selection and optimization of MS conditions. A scheduled multiple reaction monitoring (MRM) assay was designed based on the selected transitions and the retention time for further optimization, and then the acquisition method was developed for the final quantification work. For details, see SI text.

### Mass Spectrometry-Based Intermediate Analysis

Cells were cultured in the same way as described for the shake-flask fermentation. Cells were harvested 20 h after IPTG addition by immediate quenching with 15 mL quench solution (methanol: 0.5% NaCl = 2:1), and then the intermediates were extracted with 1.5 mL extraction solution (methanol:acetonitrile:water = 2:2:1) (Bennett et al., 2008). The extracts were dried using a Speedvac and dissolved in 60 μL buffer A-buffer B (4:1, v/v); the supernatant was then transferred to a clean tube for LC–MS/MS analysis with an AB SCIEX TripleTOF 5600+. For detailed information, see SI text.

## Results

### Building an In Vitro Reconstituted System for Terpenoid Biosynthesis

We chose the MVA pathway as an in vitro source of IPP and DMAPP because of the extensive biochemical information

available. This pathway involves the conversion of three equivalents of acetyl-CoA into one equivalent of IPP or DMAPP while consuming two reducing equivalents of NADPH and three equivalents of ATP (Goldstein and Brown, 1990). The production of the simple acyclic sesquiterpene farnesene was chosen as the first target for in vitro reconstitution of a terpenoid pathway. Not only because each protein component in this biosynthetic pathway can be over-expressed and purified, but also as a well-described model system it is an attractive target for improvement as a source of bio-based jet fuel (Wang et al., 2011). For in vitro reconstitution of the farnesene biosynthetic pathway, we expressed and purified nine recombinant proteins that had previously been used to assemble a farnesene synthetic pathway in *E. coli*: AtoB, ERG13, tHMG1, ERG12, ERG8, MVD1, Idi, IspA, and AFS (Fig. 1). The feasibility of the reconstituted system was confirmed by GC–MS which allowed direct detection of farnesene production from acetyl-CoA (Fig. S1). [1-$^{14}$C] acetyl-CoA was used as the substrate, resulting in a single spot by radio-thin layer chromatography (TLC) (Fig. S1), thereby allowing the convenient and rapid quantitation of farnesene production by liquid scintillation assay (Fig. 2A).

We first investigated the dependence of the reconstituted system on the concentrations of substrate, cofactors, and metal ions. The levels of farnesenemodestly increased when the concentration of acetyl-CoA was increased from 0.2 to 1 mM. Because further increases in acetyl-CoA concentration to 1.5 mM had no further benefit (Fig. 2B), the acetyl-CoA concentration was fixed at 1 mM, the same concentration of acetyl-CoA reported as intracellular levels in *E. coli* (Bennett et al., 2009). The concentrations of ATP and NADPH were independently varied from 1 to 5 mM and 0.2 to 1.5 mM, respectively, with no effect on farnesene productivity (Fig. 2 C and D). Several enzymes in this pathway have been reported to require divalent or monovalent metal ions for activity. Although$Zn^{2+}$ is known to be necessary for the activity of Idi (Carrigan and Poulter, 2003), addition of $Zn^{2+}$ to the

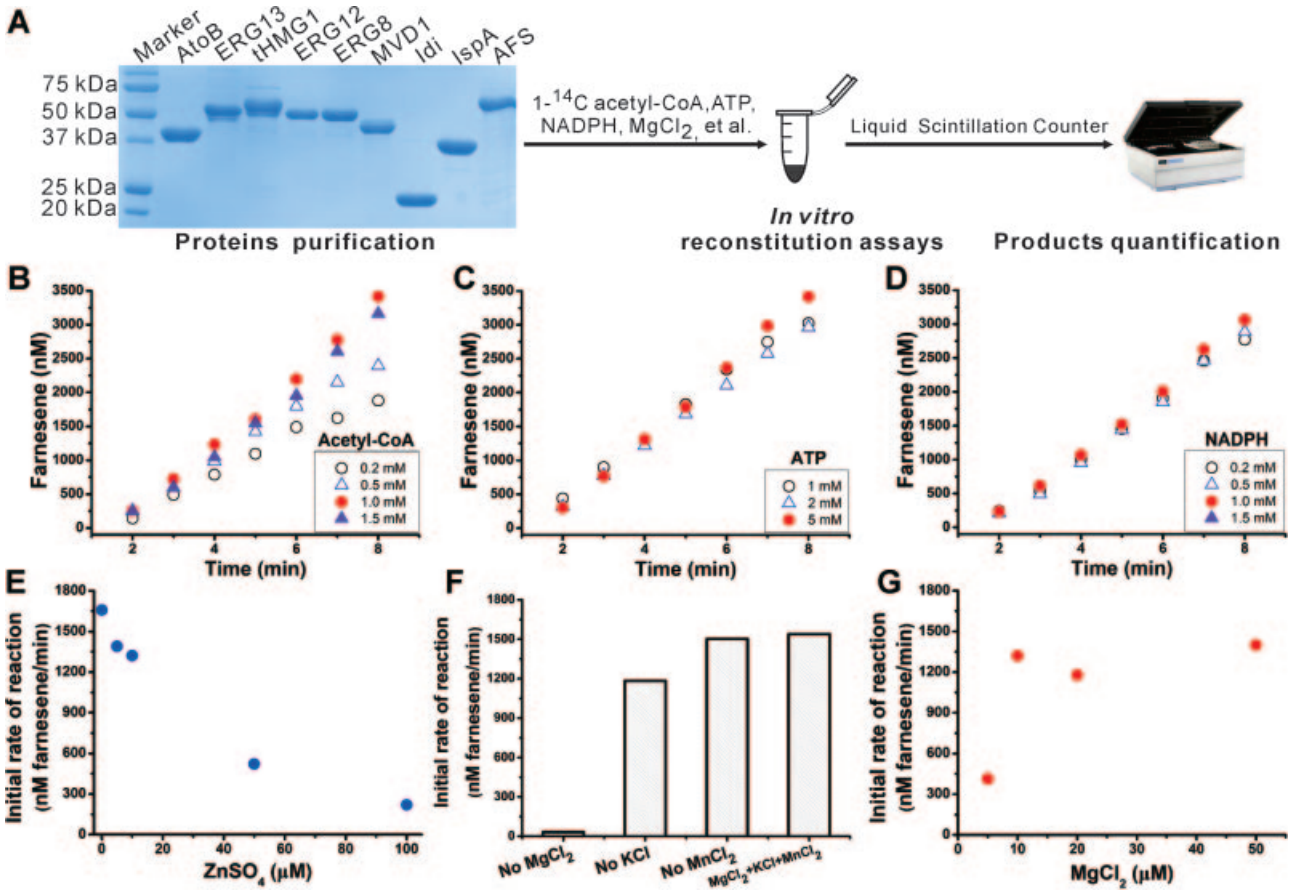


**Figure 2.** Titration of substrate, cofactors, and metals: (**A**) overview of the radio-labeled kinetic assay process, (**B**) acetyl-CoA, (**C**) ATP, (**D**) NADPH, (**E**) $ZnSO_4$, (**F**) metal dependence and (**G**) $MgCl_2$. The reaction conditions were: 0.5 μM of each protein, B–D included 1* metal ions solution; E–G included 1 mM acetyl-CoA, 5 mM ATP, 1.2 mM NADPH, and (**B**) 1.2 mM NADPH, 4 mM ATP, and varying concentrations of acetyl-CoA; (**C**) 1 mM acetyl-CoA, 1.2 mM NADPH, and varying concentrations of ATP; (**D**) 1 mM acetyl-CoA, 5 mM ATP, and varying concentrations of NADPH; (**E**) 30 mM KCl, 0.2 mM $MnCl_2$, 10 mM$MgCl_2$, and varying concentrations of $ZnSO_4$. (**F**) 30 mM KCl, 0.2 mM $MnCl_2$, and 10 mM $MgCl_2$ was added when included. (**G**) 30 mM KCl, 0.2 mM $MnCl_2$, 10 μM $ZnSO_4$, and varying concentrations of $MgCl_2$. 1* metal ions solution includes 30 mM KCl, 8 mM $MgCl_2$, 0.2 mM $MnCl_2$, and 10 μM $ZnSO_4$.

reaction resulted in a decrease in farnesene levels (Fig. 2E). Addition of $Mn^{2+}$ had no benefit, while 30 mM $K^{+}$ slightly enhanced the product yield (Fig. 2F). Of greatest importance was the level of $Mg^{2+}$, which gave a maximum rate of formation of farnesene at concentrations above 10 mM (Fig. 2G). Based on these results, an optimized buffer containing 50 mM sodium phosphate (pH 7.4), 30 mM potassium chloride, 10 mM magnesium chloride, and 1 mMTCEP, containing 1 mM NADPH, 5 mM ATP, and 1 mM acetyl-CoA (with 5% [1-$^{14}$C] acetyl-CoA) was used in subsequent experiments.

### Contributions of Each Enzyme and Optimization of the System

We examined the effect of the concentration of individual proteins on the steady-state activity of this reconstituted system by varying the concentration of each protein while fixing the concentration of the other eight enzymes at 0.5 μM. The protein constituents fell into three categories. Variation of the concentration of AtoB, tHMG1, ERG12, or MVD1 from 0.1 to 5 μM did not significantly influence the rate of farnesene synthesis (Fig. 3A and B). For the second group of proteins, increasing concentrations of ERG13, ERG8, IspA, or AFS from 0.1 to 5 μM gave a1.5–2.7-fold increase in the rate of farnesene formation, with saturation at protein concentrations higher than 0.5–2 μM, depending on the protein (Fig. 3C and D). And most notably, increases in the concentrations of the Idi, which controls the formation of DMAPP, from 0.1 to 5 μM resulted in up to a 17-fold increase in the initial rate of farnesene formation (Fig. 3E). Because it had been reported that the chimeric protein ALF, in which IspA was fused with AFS, could improve farnesene production in vivo (Wang et al., 2011), we also examined the effect of purified ALF on the in vitro system. Although replacement of IspA and AFS with ALF showed a linear dependence on ALF concentration, the maximum rates of farnesene formation did not match those observed with even low concentrations of the individual IspA and AFS proteins (Fig. 3D). In contrast to our earlier observations with the reconstituted fatty acid synthase system, we did not observe any significant inhibition by individual proteins in the in vitro farnesene synthetic system (Liu et al., 2010; Yu et al., 2011).

We next looked for possible synergistic effects of two or more protein components in the in vitro farnesene biosynthetic system. First, an orthogonal experiment was performed, with results similar to the individual titrations; Idi had the largest contribution to the system followed by ERG13 and IspA. Based on these results, ERG13, ERG8, AFS, and Idi were chosen for combinatorial manipulation. In this series of experiments, the concentrations of ERG13 (5 μM), ERG8 (2.5 μM), AFS (1 μM), and Idi (2.5 μM) were initially set at their previously established optima, while all other protein components were held at 0.5 μM. In the first set of combinatorial experiments, we examined pairwise variation of ERG13, ERG8, and AFS, which resulted in small improvements in the rate of farnesene formation (Fig. 3F).

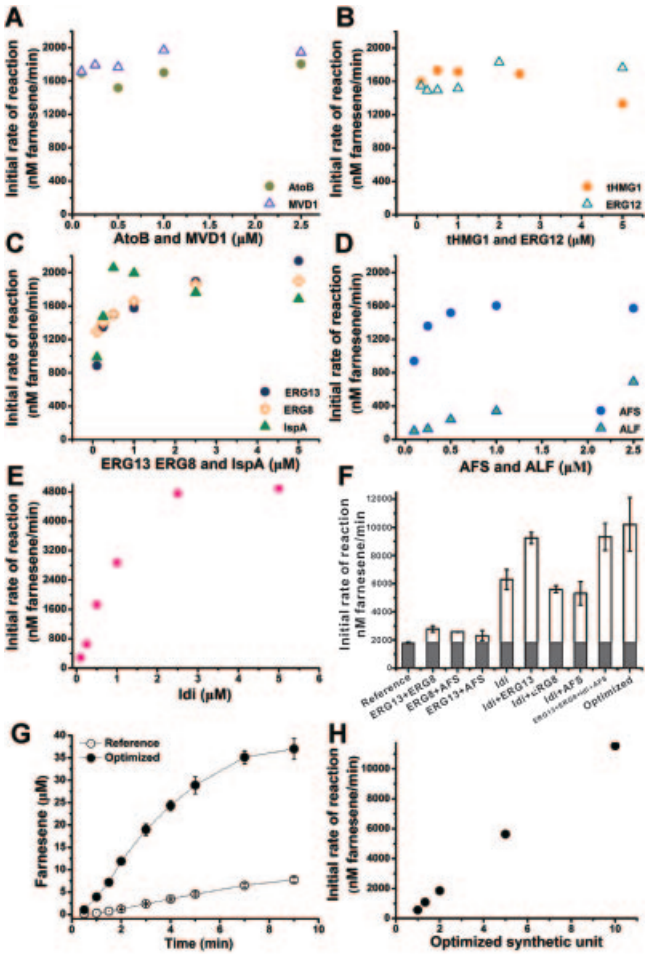


**Figure 3.** Titration of farnesene biosynthetic components and optimization of the in vitro reconstitution system: (**A**) AtoB and MVD1, (**B**) tHMG1 and ERG12, (**C**) ERG13, ERG8 and IspA, (**D**) AFS and ALF, (E) Idi, (**F**) Optimization of the in vitro reconstitution system by adjustment of the components, (**G**) Comparison of farnesene production under reference and optimized conditions, (**H**) Concentration dependence of the optimized farnesene synthetic unit. Each assay mixture included 0.5 μM of all protein components except the one being titrated. Detailed information on protein concentrations in F is summarized in suppl. Table III. Molar ratio of proteins of (H) is AtoB: ERG13: tHMG1: ERG12: ERG8: MVD1: Idi: IspA: AFS = 1:10:2:5:5:2:5:2:2; one unit protein indicates 0.05 μM AtoB and corresponding concentrations of the other proteins. The reactions were performed in 50 mM phosphate, 30 mM KCl, 10 mM $MgCl_2$, and 1 mM TCEP containing 1 mM acetyl-CoA, 5 mM ATP, and 1 mM NADPH.

In the second combinatorial set, the concentrations of Idi paired with each of the other three proteins, individually or together, were varied. While increase in the concentration of Idi alone improved the observed reaction rate up to 3.7-fold, Idi plus ERG13 or Idi plus ERG13, ERG8, and AFS together gave more than a fivefold increase in farnesene formation compared with the reference incubation. Increase in either Idi plus ERG8 or Idi plus AFS did not improve the reaction rate compared to increasing the concentration of Idi alone (Fig. 3F). It is clear that Idi plays a key role in the mevalonate pathway, and Idi plus ERG13 can efficiently improve the productivity of the reconstituted system in vitro. Based on the results of these titration studies, the molar ratio

of AtoB:ERG13:tHMG1:ERG12:ERG8:MVD1:Idi:IspA:AFS was slightly adjusted to 1:10:2:5:5:2:5:2:2 (Table SIII). Under these conditions, a sixfold enhancement in farnesene production was obtained compared to the baseline protein ratios (Fig. 3G); the maximum turnover rate of the reconstituted in vitro farnesene biosynthetic pathway approached 0.4 $s^{-1}$. In addition, while keeping these relative optimized protein ratios constant, we also varied the total protein concentration of A–B from 0.05 to 0.5 μM to determine the dependence of farnesene production on the concentration of the entire optimized biosynthetic machinery. We observed a nearly linear increase in farnesene productivity over the full range of protein concentrations examined (Fig. 3H).

### Targeted Engineering for Farnesene Overproduction

Based on the information developed from the reconstituted in vitro farnesene biosynthetic system, we constructed a rationally designed recombinant *E. coli* that could be engineered for high-level production of farnesene using *E. coli* BL21(DE3) as the host strain. The control strain *E. coli* F0 harbored only the *afs* gene on pFZ35, in which the resultant farnesene synthase relied on the endogenous *E. coli* MEP and FPP biosynthetic pathways to produce farnesene. Additional mutant hosts for farnesene production were created by introduction of the two MVA-IPP-DMAPP plasmids, pMH1, and pFZ81, along with a third pET21a(+)-derived plasmid carrying *ispA* and *afs* and varying combinations of upstream genes (Fig. 4A). The baseline strain F1 carried the full farnesene biosynthetic pathway and produced 130 mg/L of farnesene, a 400-fold increase in farnesene over the control strain F0 carrying only the heterologous *afs* gene (Fig. 4B). To further enhance the production of farnesene, strain F1 was subjected to a round of targeted in vivo engineering based on varying individual protein elements. As summarized in Figure 4A, the intracellular concentrations of ERG13, ERG8, and Idi were individually increased to generate strains F2, F3, and F4, respectively. Strain F4 exhibited an increase in farnesene production to731 mg/L 48 h after induction, representing a 5.5-fold increase over the baseline strain F1 and a 2,000-fold increase over the control strain F0 (Fig. 4B). This also represented a ~2-fold higher level of farnesene production than previous reported under the same conditions (Wang et al., 2011). Farnesene production in strain F4 reached 1.1 g/L at 96 h post induction (Fig. 4C). By contrast strains F2 and F3 showed a significant decrease in both growth and farnesene production (Fig. 4B), so further analysis was performed to confirm the status of each engineered strain. The individual relative concentrations of all of nine biosynthetic proteins were simultaneously analyzed by targeted proteomics using a modification of previously described methods to confirm the mutants work as predicted (Redding-Johanson et al., 2011). To build a highly efficient terpenoid pathway in microbial hosts, the full pathway must be reconstructed and overexpressed without metabolic bottlenecks, and the flux through the upstream and downstream modules must be carefully balanced to avoid

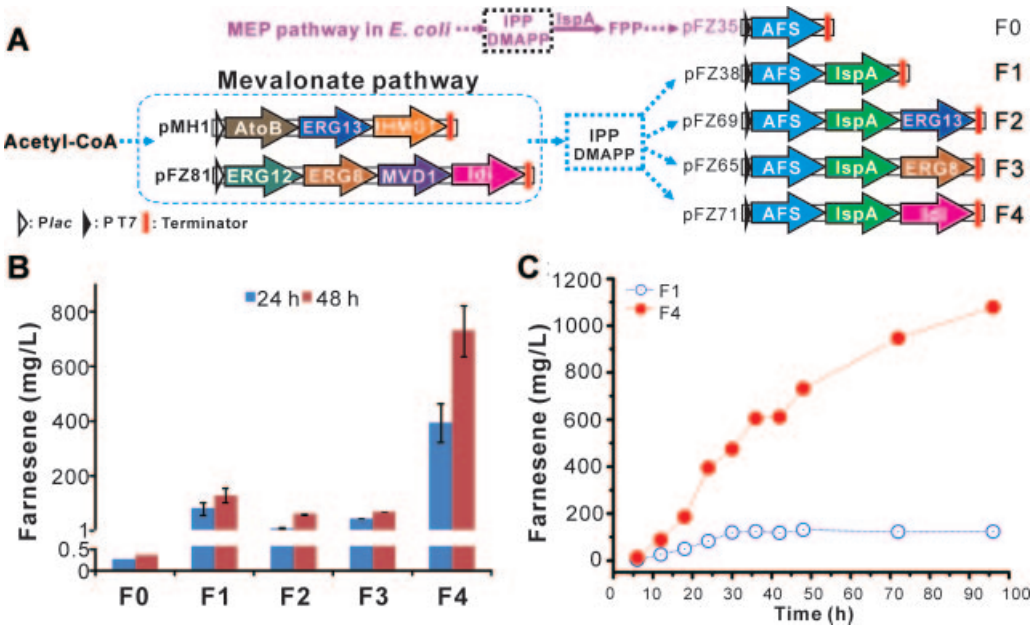


**Figure 4.** Overproduction of farnesene in *E. coli*: (**A**) Production of farnesene via the endogenous MEP or synthetic mevalonate pathways and depiction of the synthetic operons used in this study. Each operon was controlled by the indicated single promoter; plasmids pMH1 and pFZ81 were controlled by the $P_{lac}$ promoter and pET21a(+)-derived constructs were controlled by the $P_{T7}$ promoter. (**B**) Production of farnesene in engineered strains. The decane phase (1 mL) was sampled at 24 and 48 h after IPTG induction. After centrifugation, the diluted decane phase was subjected to GC–MS analysis. The quantitation of farnesene was determined from the average of three independent experiments. (**C**) Shake-flask production of farnesene from F1 and F4; the decane phase (1 mL) was sampled at the indicated time after induction by IPTG.

accumulation of toxic intermediate metabolites or diversion of feedstock to unproductive metabolism (Ajikumar et al., 2010). Therefore, key intermediates in the mutants were monitored using a MS-based relative quantification method. We observed that the level of ERG13 was increased in strain F2, as expected (Table I), and that a large amount of HMG-CoA accumulated (Table II) due to ERG13 overexpression. This is consistent with previous reports showing that accumulation of HMG-CoA can inhibit fatty acid biosynthesis (Kizer et al., 2008). In strain F3, there was no significant benefit reflected in farnesene production (Fig. 4B) and too much ERG8 (hundreds of folds) did not benefit the system (Table I), as also observed in the corresponding in vitro assays (Fig. 3C), it maybe attribute to the IPP/DMAPP accumulation (Table II) which can inhibit cell growth (Martin et al., 2003). Interestingly, strain F4 works much smoother than the other strains because fewer inhibitors (HMG-CoA, IPP/DMAPP, and FPP) accumulated in F4 than other strains, but the concentration of acetyl-CoA, ATP, and NADPH was also lower than the others (Table II), with an increase in levels of Idi (Table I). Based on the observation that the levels of overexpression of one protein in a biosynthetic pathway influences the expression of the remaining pathway proteins (Table I), further enhancement of farnesene production levels by target strain engineering will require the development of new methods for quantitative control of each biosynthetic protein.

## Discussion

During the past 10 years, the production of terpenoid metabolites in bacteria has achieved notable initial success, not only for small molecules (Alper et al., 2005; Peralta-Yahya et al., 2011, 2012), but also for complex drug precursor compounds (Ajikumar et al., 2010; Tsuruta et al., 2009). These achievements, however, were based on traditional labor-intensive combinations of molecular cloning or high throughput screening method (Dietrich et al., 2010; Wang et al., 2009). Despite the availability of improved mutants, further improvement has been difficult due to lack of information about steady-state kinetics or the influence of protein concentrations and metabolic flux on the efficiency of terpenoid biosynthetic pathways. In many cases it is evident that the optimal conditions for a multistep pathway represent a very narrow range, with identification of the actual optimum balance inaccessible solely by random recombination or mutation (Ajikumar et al., 2010).

Here, we have fully reconstituted an in vitro synthetic pathway for farnesene production from individually purified components using acetyl-CoA as the substrate for the first time. Engineering of bacterial terpene synthesis based on the MVA pathway has a number of advantages relative to the previously described system for in vitro reconstitution of fatty acid biosynthesis (Yu et al., 2011). Importantly, in the MVA system there is no inhibition of the overall pathway associated with the titer of a single protein component. Moreover, farnesene productivity is nearly linearly dependent on the titer of the optimized synthetic assembly. By using targeted proteomic analysis of engineered cultures to quantify the in vivo relative expression level of each biosynthetic protein component had provided guidelines for further pathway and strain engineering. As a proof of principle, we have demonstrated that, with a relatively small number of data-guided steps, it has been possible to engineer *E. coli* strains for production of farnesene at relatively high-levels. We expect that this target engineering strategy is general and can be readily extended to the bacterial production of any desired

**Table I.** Relative protein concentrations in engineered strains for farnesene production.

| Proteins | AtoB | ERG13 | tHMG1 | ERG12 | ERG8 | MVD1 | Idi | IspA | AFS |
|---|---|---|---|---|---|---|---|---|---|
| F1 | 40 ± 1.5 | 65 ± 3.7 | 8.4 ± 0.7 | 6.3 ± 0.8 | 4.0 ± 0.2 | 78 ± 0.1 | 16 ± 2.5 | 102 ± 3.9 | 43 ± 1.2 |
| F2 | 187 ± 14.9 | **440 ± 1.9** | 143 ± 16.3 | 33 ± 5.0 | 17 ± 1.2 | 738 ± 129 | 25 ± 0.1 | 70 ± 3.6 | 42 ± 1.5 |
| F3 | 119 ± 1.0 | 178 ± 5.5 | 10 ± 1.3 | 11 ± 0.8 | **1080 ± 55** | 275 ± 23.5 | 25 ± 0.9 | 7.0 ± 1.1 | 34 ± 0.8 |
| F4 | 90 ± 17.3 | 3.2 ± 0.6 | 48 ± 0.1 | 3.5 ± 2.1 | 1.6 ± 0.1 | 3.7 ± 1.3 | **967 ± 35** | 64 ± 3.0 | 5.2 ± 0.1 |

Cells were resuspended in 100 mM $NH_4HCO_3$ and lysed by sonicator, after ultracentrifugation 100 μg total protein was used for targeted proteomics analysis. The trypsin digested peptides were analyzed by 4500Q-Trap LC–MS/MS.

Data present the relative protein concentration (normalized average peak area/corresponding protein molecular weight × 1,000,000) ± SD of four replicates. The information of selected peptides used are listed in Table SIV.

The bold number means that the protein concentration was overexpressed as expected.

**Table II.** Intermediates accumulation in engineered strains.

| | Acetyl-CoA | HMG-CoA | MVA | IPP/DMAPP | FPP | CoA | ATP | ADP | NADPH | $NADP^+$ |
|---|---|---|---|---|---|---|---|---|---|---|
| F1 | 140 ± 29 | 10.3 ± 2.6 | 1425 ± 58 | 1272 ± 190 | 201 ± 36 | 19 ± 1.2 | 109 ± 15 | 587 ± 79 | 3.3 ± 0.4 | 97 ± 9 |
| F2 | 574 ± 56 | **235 ± 52** | 1468 ± 39 | 133 ± 21 | 91 ± 30 | 148 ± 15 | 171 ± 36 | 434 ± 98 | 7.0 ± 1.6 | 195 ± 45 |
| F3 | 90 ± 7 | 14 ± 2.3 | 1240 ± 63 | **2138 ± 42** | 128 ± 40 | 9.3 ± 1.6 | 106 ± 6 | 452 ± 12 | 5.5 ± 1.0 | 168 ± 12 |
| F4 | **20 ± 4.9** | 24 ± 6 | 745 ± 29 | **3.0 ± 0.6** | **12 ± 0.6** | 2.9 ± 0.5 | **14 ± 2.8** | 112 ± 9 | **1.6 ± 0.4** | 84 ± 3.0 |

Cells were quenched with methanol: 0.9% NaCl (2:1, v/v) and intermediates were extracted with 1.5 mL methanol: acetonitrile:water (2:2:1, v/v). The extracts were analyzed by TripleTOF 5600+ high resolution LC–MS/MS.

Data present the (average peak area ± SD)/10,000 of the intermediates of four replicates.

The bold number means that the concentration was significantly changed compared with other strains.

terpenoid. By using this strategy, identification of those enzymes that are key enzymes of the indicated pathway is much easier. This approach can also be an excellent supplement to currently used randomized high throughput methods for generation of pathways and targeted screening by decreasing variables and providing guidelines for further engineering. As we know, screening around the key components of a pathway under an optimal conditions can significantly simplify the approach, it can promote the high throughput method to be a more purposeful and economical strategy.

There are, however, limitations to the in vitro reconstituted system. First, all necessary enzymes must be overexpressed and purified with high activity to mimic in vivo conditions, so it is not suitable for the pathway contains protein overexpressed without desired activity. Second, the accumulation of synthetic pathway intermediates or side reactions of individual overexpressed components may influence other in vivo metabolic pathways resulting in systemic imbalance, and this cannot be addressed by the in vitro system alone, such as the effect of accumulation of HMG-CoA on fatty acid biosynthesis (Kizer et al., 2008) or inhibition of cell growth by elevated levels of IPP and DMAPP (Martin et al., 2003). Nonetheless, careful optimization of the in vitro system allows identification and quantification of the accumulation of any pathway intermediate by MS, and a protein scaffold or a dynamic stress–response promoter strategy can be employed in vivo to avoid accumulation of toxic intermediates (Dahl et al., 2013; Dueber et al., 2009). The fact that an optimized strain such as F4 exhibits healthy growth and accumulation of fewer intermediates (Table II) suggests that the optimized pathway is intrinsically in balance. The concentration of acetyl-CoA, ATP, and NADPH in the optimized strain is also less in F4 than in the other strains (Table II). This suggests further engineering work could be done to supply more acetyl-CoA, ATP, and NADPH, such as engineering the TCA cycle, pentose phosphate pathway, or Entner–Doudoroff pathway (He et al., 2013; Zhao et al., 2013). Third, although cofactor dependence can be titrated in vitro, an in vitro experiment cannot achieve the same constant ratio of ATP to ADP and NADPH to $NADP^+$ as in living cells. The in vitro assays also cannot reflect the effect of high cell maintenance or membrane stress from product accumulation (on cellular product synthesis). Metabolic flux analysis (13C-MFA) can be employed to quantify metabolism and demonstrate its response to the engineered work, especially for the cofactors and energy balance analyses (He et al., 2013). Fourth, our in vitro method cannot reflect the natural substrate channeling between enzymes in side of the cell. Lastly, the current techniques for precise expression of multiple proteins is limited, so creating a pathway comprised of precisely expressed proteins is not possible.

Analysis of the reconstituted in vitro system has enabled us to quantify its steady-state kinetic parameters and providing a blueprint for the subsequent in vivo optimization of the farnesene biosynthetic pathway. However, the targeted proteomic analysis of protein concentration has shown that overexpression of one protein can influence the levels of expression of other proteins in the pathway (Table I), so future targeted engineering still requires new methods to quantitatively control the titers of all protein components in vivo.

We thank Prof. David E. Cane, Prof. Sir. David Hopwood and Prof. Chaitan Khosla for assistance in the preparation of the manuscript and for critical suggestions, and Shuai Fu from J1 Biotech and Dr. Jianlan Yu from AB Sciex for helping with proteomics and intermediates analysis experiments. This research was supported by grants from the 973 (2011CBA00800, 2012CB721000) and 863 (2012AA02A701) Project from the Ministry of Science and Technology of China and, the National Natural Science Foundation of China (31222002) as well as Program for New Century Excellent Talents in University to T. Liu and Doctoral Fund of Ministry of Education of China (20110141120028).

## References

Ajikumar PK, Tyo K, Carlsen S, Mucha O, Phon TH, Stephanopoulos G. 2008. Terpenoids: Opportunities for biosynthesis of natural product drugs using engineered microorganisms. Mol Pharm 5(2):167–190.

Ajikumar PK, Xiao WH, Tyo KEJ, Wang Y, Simeon F, Leonard E, Mucha O, Phon TH, Pfeifer B, Stephanopoulos G. 2010. Isoprenoid pathway optimization for Taxol precursor overproduction in *Escherichia coli*. Science 330(6000):70–74.

Alper H, Miyaoku K, Stephanopoulos G. 2005. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23(5):612–616.

Bennett BD, Kimball EH, Gao M, Osterhout R, Van Dien SJ, Rabinowitz JD. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. Nat Chem Biol 5(8):593–599.

Bennett BD, Yuan J, Kimball EH, Rabinowitz JD. 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat Protoc 3(8):1299–1311.

Carrigan CN, Poulter CD. 2003. Zinc is an essential cofactor for type I isopentenyl diphosphate: Dimethylallyl diphosphate isomerase. J Am Chem Soc 125(30):9008–9009.

Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD. 2013. Engineering dynamic pathway regulation using stress–response promoters. Nat Biotechnol 31:1039–1046.

Dietrich JA, McKee AE, Keasling JD. 2010. High-throughput metabolic engineering: Advances in small-molecule screening and selection. Annu Rev Biochem 79:563–590.

Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KLJ, Keasling JD. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27(8):753–759.

Farmer WR, Liao JC. 2000. Improving lycopene production in *Escherichia coli* by engineering metabolic control. Nat Biotechnol 18(5):533–537.

Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349–361.

Goldstein JL, Brown MS. 1990. Regulation of the mevalonate pathway. Nature 343(6257):425–430.

He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, Peng L. 2014. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on 13C-metabolic flux analysis. Biotechnol Bioeng 111(3):575–585.

Kirby J, Keasling JD. 2009. Biosynthesis of plant isoprenoids: Perspectives for microbial engineering. Annu Rev Plant Biol 60:335–355.

Kizer L, Pitera DJ, Pfleger BF, Keasling JD. 2008. Application of functional genomics to pathway optimization for increased isoprenoid production. Appl Environ Microbiol 74(10):3229–3241.

Liu T, Vora H, Khosla C. 2010. Quantitative analysis and engineering of fatty acid biosynthesis in *E. coli*. Metab Eng 12(4):378–386.

Martin VJJ, Pitera DJ, Withers ST, Newman JD, Keasling JD. 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat Biotechnol 21(7):796–802.

Paddon C, Westfall P, Pitera D, Benjamin K, Fisher K, McPhee D, Leavell M, Tai A, Main A, Eng D. 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496:528–532.

Pechous SW, Whitaker BD. 2004. Cloning and functional expression of an (E, E)-a-farnesene synthase cDNA from peel tissue of apple fruit. Planta 219(1):84–94.

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS. 2011. Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483.

Peralta-Yahya PP, Zhang F, del Cardayre SB, Keasling JD. 2012. Microbial engineering for the production of advanced biofuels. Nature 488(7411):320–328.

Phillips MA, León P, Boronat A, Rodríguez-Concepción M. 2008. The plastidial MEP pathway: Unified nomenclature and resources. Trends Plant Sci 13(12):619–623.

Polakowski T, Stahl U, Lang C. 1998. Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. Appl Microbiol Biotechnol 49(1):66–71.

Redding-Johanson AM, Batth TS, Chan R, Krupa R, Szmidt HL, Adams PD, Keasling JD, Soon Lee T, Mukhopadhyay A, Petzold CJ. 2011. Targeted proteomics for metabolic pathway optimization: Application to terpene production. Metab Eng 13(2):194–203.

Rohmer M, Knani M, Simonin P, Sutter B, Sahm H. 1993. Isoprenoid biosynthesis in bacteria: A novel pathway for the early steps leading to isopentenyl diphosphate. Biochem J 295:517–524.

Shao Z, Zhao H. 2009. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucleic Acids Res 37(2):e16.

Singh P, Batth TS, Juminaga D, Dahl RH, Keasling JD, Adams PD, Petzold CJ. 2012. Application of targeted proteomics to metabolically engineered *Escherichia coli*. Proteomics 12(8):1289–1299.

Tsuruta H, Paddon CJ, Eng D, Lenihan JR, Horning T, Anthony LC, Regentin R, Keasling JD, Renninger NS, Newman JD. 2009. High-level production of amorpha-4, 11-diene, a precursor of the antimalarial agent artemisinin, in *Escherichia coli*. PloS ONE 4(2):e4489.

Unwin RD, Griffiths JR, Whetton AD. 2009. A sensitive mass spectrometric method for hypothesis-driven detection of peptide post-translational modifications: Multiple reaction monitoring-initiated detection and sequencing (MIDAS). Nat Protoc 4(6):870–877.

Wang C, Yoon SH, Jang HJ, Chung YR, Kim JY, Choi ES, Kim SW. 2011. Metabolic engineering of *Escherichia coli* for α-farnesene production. Metab Eng 13:648–655.

Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, Forest CR, Church GM. 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460(7257):894–898.

Westfall PJ, Pitera DJ, Lenihan JR, Eng D, Woolard FX, Regentin R, Horning T, Tsuruta H, Melis DJ, Owens A. 2012. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. Proc Natl Acad Sci USA 109(3):E111–E118.

You C, Zhang XZ, Zhang YHP. 2012. Simple cloning via direct transformation of PCR product (DNA multimer) to *Escherichia coli* and *Bacillus subtilis*. Appl Environ Microbiol 78(5):1593–1595.

Yu X, Liu T, Zhu F, Khosla C. 2011. In vitro reconstitution and steady-state analysis of the fatty acid synthase from *Escherichia coli*. Proc Natl Acad Sci USA 108(46):18643–18648.

Zhang F, Rodriguez S, Keasling JD. 2011. Metabolic engineering of microbial pathways for advanced biofuels production. Curr Opin Biotechnol 22(6):775–783.

Zhao J, Li Q, Sun T, Zhu X, Xu H, Tang J, Zhang X, Ma Y. 2013. Engineering central metabolic modules of *Escherichia coli* for improving β-carotene production. Metab Eng 17:42–50.

## Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web-site.